HiEve: A Corpus for Extracting Event Hierarchies from News Stories

Goran Glavaš[1], Jan Šnajder[1], Parisa Kordjamshidi[2], and Marie-Francine Moens[2]

[1]University of Zagreb, Faculty of Electrical Engineering and Computing
Text Analysis and Knowledge Engineering Lab
Unska 3, 10000 Zagreb, Croatia
{goran.glavas, jan.snajder}@fer.hr

[2]KU Leuven, Department of Computer Science
Language Intelligence & Information Retrieval Group
Celestijnenlaan 200A, B-3000 Leuven, Belgium
{parisa.kordjamshidi, sien.moens}@cs.kuleuven.be

Abstract

In news stories, event mentions denote real-world events of different spatial and temporal granularity. Narratives in news stories typically describe some real-world event of coarse spatial and temporal granularity along with its subevents. In this work, we present HiEve, a corpus for recognizing relations of spatiotemporal containment between events. In HiEve, the narratives are represented as hierarchies of events based on relations of spatiotemporal containment (i.e., *superevent–subevent* relations). We describe the process of manual annotation of HiEve. Furthermore, we build a supervised classifier for recognizing spatiotemporal containment between events to serve as a baseline for future research. Preliminary experimental results are encouraging, with classifier performance reaching 58% F1-score, only 11% less than the inter-annotator agreement.

Keywords: Event hierarchies, spatiotemporal containment, relation extraction

1. Introduction and Motivation

In natural language processing (NLP), events are often defined rather vaguely as things that *happen* or *occur* (Pustejovsky et al., 2003a). Although events in the real-world have precise (and also unique) spatiotemporal extent (Quine, 1985), their linguistic counterparts, *event mentions*, are often spatially and temporally rather vague or imprecise (Glavaš and Šnajder, 2013b). Considering this common absence of precise spatiotemporal locations of events, NLP tasks aiming to extract spatial or temporal properties of events have been proven rather difficult (Roberts et al., 2013; UzZaman et al., 2013; Glavaš and Šnajder, 2013a).

Event mentions in news stories are, arguably, particularly spatiotemporally vague, because news stories put more focus on structure and coherence of the narrative (Wolf and Gibson, 2005) than on spatiotemporal properties of the individual events. Furthermore, different event mentions often denote events from the real-world with varying levels of granularity (e.g., a *championship* is coarser than a *match*, which is, in turn, coarser than *scoring of a goal*). Not addressing the issue of granularity further debilitates the inference over event mentions.

Previous research represented narratives in news as chains of individual events in which one event strictly precedes the other (Chambers and Jurafsky, 2009; Chambers and Jurafsky, 2010; Jans et al., 2012). However, such linear representation is not expressive enough because (1) events of the same narrative may be in temporal relations other than precedence (i.e., temporal containment or equivalence), (2) events of the same narrative are not only temporally but also spatially related, and (3) event mentions are of varying granularity and some events constitute a part of some other events.

Considering that a single narrative in news relates to a single real-world event with all its constituting subevents, in this work we propose a hierarchically structured representation of news narratives. More precisely, we represent news stories as *event hierarchies* – directed acyclic graphs (DAGs) of event mentions in which edges denote spatiotemporal containment between events. The relation of spatiotemporal containment indicates that one event (to which we refer as *subevent*) is a part of the other event (to which we refer as *superevent*). For an event to be a part of another event, it needs to be both spatially and temporally contained in the other event. Neither spatial nor temporal dimension alone are enough to infer the *part-of* relation (i.e., the spatiotemporal containment) between events. More concretely, neither spatial containment (e.g., CONTAINS relation in SpatialML) nor temporal containment (e.g., DURING relation in TimeML) alone imply that one event is a part of another. Temporally, an event may occur DURING another event but not be a part of it. E.g., in example (1), the *"revolution"* happened during the *"World War II"*, but was not a part of the *"War"* because the *"revolution"* took place in Argentina, whereas the World War II was not led on Argentinian soil. On the other hand, *"reduction of rents"*, *"lowering of taxes"*, and *"making hospitals free"* happened during the *"revolution"*, but were all part of the *"revolution"* as they were happening at the same location, in Argentina.

> (1) *During the* ***revolution*** *in '43, in the midst of the World* ***War*** *II, the new Argentinian government* ***reduced*** *rents,* ***lowered*** *taxes, and* ***made*** *hospitals free.*

Similarly, one event may be spatially CONTAINED by the other event but not be a part of that event. In example (2), the *"plague"* ravaged the entire city of London, whereas the fire *"destroyed"* 60% of the city. However, the *"destroyed"* is not a part of the *"plague"* because it is not temporally contained by the event *"plague"* (it happened afterwards).

(2) *The fire **destroyed** 60% of London after almost 30,000 people died from **plague**.*

In this paper, we present a corpus of manually annotated, spatiotemporally coherent event hierarchies in news stories. The purpose of such a resource is to enable supervised extraction of event-based document representations that jointly models spatial and temporal containment between events. Such a representation would enable joint spatiotemporal reasoning over events, which would be useful for a number of NLP applications, e.g., event-oriented information retrieval (Glavaš and Šnajder, 2013a) or event-based document summarization (Daniel et al., 2003; Filatova and Hatzivassiloglou, 2004). We also present a baseline supervised model for pairwise recognition of spatiotemporal containment between event mentions. The performance of the proposed baseline model is encouraging, suggesting that the extraction of event hierarchies from text is feasible.

2. Related Work

The emergence of the TimeML standard (Pustejovsky et al., 2003a) and the TimeML abiding TimeBank corpus (Pustejovsky et al., 2003b) triggered a large body of work on extracting events and temporal information from text. Much of this work was carried out within designated evaluation campaigns such as Automated Content Extraction (ACE) (ACE, 2005; ACE, 2007) and TempEval (Verhagen et al., 2007; Verhagen et al., 2010; UzZaman et al., 2013). All three TempEval evaluations featured at least one task on extracting temporal relations between events. In TempEval-2 (Verhagen et al., 2010), the best performance was achieved by the system using supervised machine learning with Markov logic networks (UzZaman and Allen, 2010). In the TempEval-3 (UzZaman et al., 2013), the best performing system on temporal relation extraction task (Bethard, 2013) employed the linear SVM and logistic regression in combination with a set of simple morpho-syntactic features.

More recently, following the emergence of standards for annotating spatial relations (Mani et al., 2008; Kordjamshidi et al., 2010), Roberts et al. proposed an annotation scheme (2012) and developed a supervised approach for extracting spatial relations between events (2013). They use the linear SVM classifier on the feature set similar to the features sets used for temporal relation extraction to distinguish between six types of spatial relations. The achieved performance on spatial relation classification of 60% F-score is comparable to the state-of-the-art performance on temporal relation extraction. As argued in the Introduction, neither spatial nor temporal containment alone, however, suffice to indicate that one event is part of the other event. In this work we focus on spatiotemporal containment between events, i.e., relations that indicate that one event constitutes the other.

Chambers and Jurafsky (2008) represent narratives in news stories as chains of temporally ordered event mentions sharing a common participant. In many cases, however, limiting a narrative to a sequence of participant-sharing events can be overly restrictive, as it does not allow for relations between events that do not share protagonists (e.g., the *"meeting between Obama and Putin"* and the *"lunch that Merkel and Berlusconi had together"* may belong to the same narrative, which, e.g., describes one particular *"G8 meeting"*). In their subsequent work, Chambers and Jurafsky (2009) go on to define *event schemas* (i.e., sets of event chains) as means of modeling all the actors participating in the set of events. Although related to our work, the extraction of event schemas differ from the extraction of spatiotemporal event hierarchies in two important aspects. First, only verbal events are considered for participation in narrative chains, whereas nominal event mentions (e.g., *"elections"*, *"murder"*, *"match"*), which can be very informationally relevant, are neglected. Second, unlike event chains, in spatiotemporal hierarchies of events we allow for relations between events that do not have any protagonists in common.

3. HiEve Corpus

As a starting point for the annotation process, we randomly selected 100 documents from the GraphEve corpus,[1] which contains gold-annotated event mentions. Event mentions annotated in the GraphEve corpus conform to the restricted TimeML definition (Pustejovsky et al., 2003a) – only factual TimeML instances are annotated (i.e., only mentions denoting events that actually happened).

3.1. Annotation Guidelines

Two annotators were given the task to independently annotate event hierarchies in news stories as DAGs denoting spatiotemporal containment between event mentions. An example of a news story snippet and its accompanying DAG is given in Fig. 1 More precisely, given a pair of event mentions, annotators were instructed to annotate one of the following relation types:

1. SUPERSUB relation, which denotes that the first event of the pair spatiotemporally contains the second event, i.e., the event denoted by the second mention is part of the event denoted by the first mention;

2. SUBSUPER relation, which denotes that the second event of the pair spatiotemporally contains the first event, i.e., the event denoted by the first mention is part of the event denoted by the second mention;

3. COREF relation, which denotes that two event mentions denote the same real-world event;

4. NORELATION annotation, which denotes that neither of the events spatiotemporally contains the other, nor do event mentions corefer.

In order to obtain consistent annotations, we provided annotators with the following set of annotation guidelines:

[1] `http://takelab.fer.hr/data/grapheve/`

U.S. President Barack Obama ***sparred*** *with Russia's Vladimir Putin over how to end the* ***war*** *in Syria on Monday during an icy* ***encounter*** *at a G8* ***summit****.* ***Speaking*** *after* ***talks*** *with Obama, Putin* ***said*** *Moscow and Washington agreed the* ***bloodshed*** *must stop...*

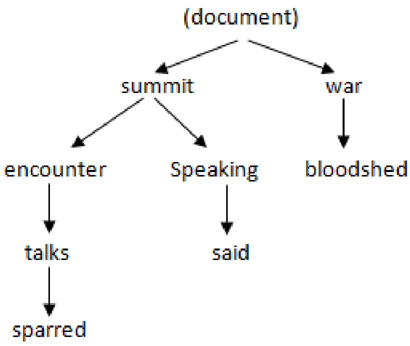

Figure 1: Example of an event hierarchy

1. Annotate pairs of events in which one spatiotemporally contains the other (e.g., *"revolution"* and *"reduced"* in example (1)). Do not annotate relations between pairs of events in which one only spatially (e.g., *"plague"* and *"destroyed"* in example (2)) or only temporally (*"War"* and *"revolution"* in example (1)) contains the other;

2. Among the events that spatiotemporally contain some event e_i, always select the one with the smallest spatiotemporal extent as the superevent of e_i. E.g., in example (3), both *"bloodshed"* and *"gunfights"* spatiotemporally contain *"shooting"*, but we annotate *"gunfights"* as the superevent of *"shooting"* because it has a smaller spatiotemporal extent than *"bloodshed"*;

3. In cases where there are several coreferent mentions of the superevent, annotate the relation between the subevent and the coreferent mention of the superevent that is closest to the subevent in the text;

4. Annotate coreference by assigning each mention to the closest previous mention that refers to the same real-world event, if such exists. The SUPERSUB and SUBSUPER relations of a mention need not be re-annotated for its coreferent mentions, as these can be inferred by employing transitivity, which emanates from SUPERSUB, SUBSUPER, and COREF relations.

(3) *Syrian forces were* ***shooting*** *in fierce* ***gunfights*** *with rebels in one city and* ***shelled*** *another on Monday, continuing a year-long* ***bloodshed****.*

3.2. Inter-Annotator Agreement

We first let both annotators independently annotate the same 30 documents, on which we compute the inter-annotator agreement (IAA). After that, each of the annotators annotated her own set of 35 documents. The time required to annotate a single document was, on average, 30 minutes.

The approach we use for measuring inter-annotator agreement is a variation of the approach proposed for temporal relations by UzZaman and Allen (2011). Let $R_{STC}(a_i)$ be the set of spatiotemporal containment relations (i.e., union of SUPERSUB and SUBSUPER relations) annotated by the annotator a_i and $closure(R)$ be the transitive closure of R derived based on transitiveness stemming from SUPERSUB, SUBSUPER, and COREF relations. The agreement can be interpreted as the F-score, i.e., the harmonic mean of precision and recall, computed as follows:

$$precision = \frac{|r \in R_{STC}(a_1) : r \in closure(R(a_2))|}{|r \in R_{STC}(a_1)|}$$

$$recall = \frac{|r \in R_{STC}(a_2) : r \in closure(R(a_1))|}{|r \in R_{STC}(a_2)|}$$

The SUPERSUB and SUBSUPER relations are introduced via transitive closure, according to the following transitivity rules:

$$e_1 \text{ COREF } e_2 \wedge e_2 \text{ COREF } e_3 \Rightarrow e_1 \text{ COREF } e_3$$
$$e_1 \text{ SUPERSUB } e_2 \wedge e_2 \text{ SUPERSUB } e_3 \Rightarrow e_1 \text{ SUPERSUB } e_3$$
$$e_1 \text{ SUBSUPER } e_2 \wedge e_2 \text{ SUBSUPER } e_3 \Rightarrow e_1 \text{ SUBSUPER } e_3$$
$$e_1 \text{ SUPERSUB } e_2 \wedge e_2 \text{ COREF } e_3 \Rightarrow e_1 \text{ SUPERSUB } e_3$$
$$e_1 \text{ SUPERSUB } e_2 \wedge e_1 \text{ COREF } e_3 \Rightarrow e_2 \text{ SUBSUPER } e_3$$
$$e_1 \text{ SUBSUPER } e_2 \wedge e_2 \text{ COREF } e_3 \Rightarrow e_1 \text{ SUBSUPER } e_3$$
$$e_1 \text{ SUBSUPER } e_2 \wedge e_1 \text{ COREF } e_3 \Rightarrow e_2 \text{ SUPERSUB } e_3$$

The overall observed IAA was 69% F-score. We believe this to be a fair agreement considering that the task at hand amounts to jointly annotating spatial and temporal containment between events. Annotators themselves judged the task as very cognitively demanding. Brief inspection of the disagreements reveals that a fair share of them originate from obvious annotation mistakes caused by the lack of concentration, confirming the intuition that annotating spatiotemporal containment between events is a cognitively demanding task. However, this also indicates that there is room for further improvements in the annotation quality. In order to consolidate the annotations for the learning process, the annotators worked together to resolve the disagreements.

3.3. Corpus Analysis

The compiled HiEve corpus consists of 100 documents (1354 sentences, 33273 tokens) containing, on average, 32 event mentions.[2] On average, annotators annotated 10 SUPERSUB and SUBSUPER relations per document, which, after applying transitive closure over SUPERSUB, SUBSUPER, and COREF relations, gave on average 23 SUPERSUB and SUBSUPER relations per document.

From a perspective of automated extraction of event hierarchies, one interesting analysis is that of distance between the event mentions constituting SUPERSUB and SUBSUPER relations. The distribution of SUPERSUB and SUBSUPER relations over distances between their respective

[2] HiEve corpus is available under the CC BY-NC-SA 3.0 license from http://takelab.fer.hr/hievents.rar

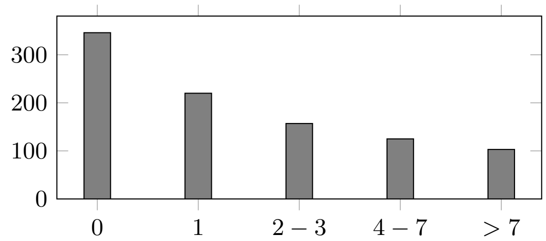

Figure 2: Histogram of distances between event mentions in spatiotemporal containment relations

event mentions may be informative regarding the type of information needed to determine the spatiotemporal containment of events (e.g., syntactic information may be helpful if most SUPERSUB and SUBSUPER relations occur between event mentions of the same sentence). The histogram of relations of spatiotemporal containment over distances between event mentions (in number of sentences) is shown in Fig. 2. Most annotated relations of spatiotemporal containment hold between events from the same sentence or adjacent sentences. Relations in which event mentions are not more than three sentences apart account for 76% of all SUPERSUB and SUBSUPER relations. However, the proportion of long-distance relations, in which the mentions are more than seven sentences apart, is not negligible (11%).

4. Pairwise Spatiotemporal Containment

We propose a supervised model for recognizing SUPERSUB and SUBSUPER relations between pairs of event mentions in order to (1) get the feeling about the difficulty of automated recognition of spatiotemporal containment between events from text and (2) set a reasonable baseline for future research.

4.1. Features

The supervised model is based on the set of features that fall into the following four feature groups:

1. **Event-based features** (EVE) include word, lemma, stem, and POS-tag of both event mentions. We use additional features that compare the arguments of event mentions of three different coarse-grained argument types: AGENT, TARGET, and LOCATION. Arguments were extracted automatically using the rule-based approach proposed by Glavaš and Šnajder (2013b);

2. **Bag-of-words features** (BoW) include all the lemmas between the two event mentions but also all the temporal prepositions (*before*, *after*, *during*, etc.) and all the spatial prepositions (*above*, *below*, *behind*, etc.) between the mentions;

3. **Positional features** (PS) indicate the distance between event mentions, both in number of sentences and number of tokens. Additionally, we use a feature indicating whether the events mentions are adjacent (no other event mentions in between);

4. **Syntactic features** (SYN) are computed only for pairs of events from the same sentence. This set includes all the relations on the path between the two mentions in the dependency tree and the features indicating whether one of the event mentions syntactically governs the other;

4.2. Classification

We cast the problem of determining pairwise spatiotemporal containment between events as a ternary classification task with the following classes: (1) first event mention spatiotemporally contains the second (SUPERSUB), (2) second event mention spatiotemporally contains the first (SUBSUPER), and (3) no spatiotemporal containment between the events (NOCONTAINMENT).

We compile the pairs of event mentions so that the first event of the pair is always the one that occurs before in the document. We take all the SUPERSUB and SUBSUPER relations (including those obtained using transitivity) as positive training examples. In order to make the evaluation realistic, we treat all other pairs of event mentions as negative examples (i.e., instances of the NOCONTAINMENT class). This way we obtain in total 29,956 pairs of event mentions, 1,112 of which SUPERSUB pairs and 1,145 SUBSUPER pairs. We then split the dataset into training and test portions (70:30 ratio).

With the number of features being much larger than the number of examples, we employ a linear discriminative model for classification, namely the L2-regularized logistic regression. We used the LibLinear (Fan et al., 2008) implementation of the logistic regression. We optimize the hyperparameters of the learning algorithm via grid search using 10-fold cross-validation on the training set. We then report the results of the optimal model on the test set. To analyze the contribution of different feature sets, in Table 1 we show the performance (in terms of F_1-score) of models employing various combinations of feature sets. The results are given separately for the SUPERSUB and SUBSUPER relations, together with the micro-averaged performance over these two classes.

The overall best performance of 58% F_1-score is encouraging, especially considering that the IAA is only 11% higher and also considering that we employed a rather simple feature set. Results suggest that event-based (EVE) and bag-of-words (BoW) features contribute the most to recognizing spatiotemporal containment between events. Expectedly, PS features are poor predictors on their own, but contribute to better recognition of SUBSUPER relations when combined with EVE and BoW features. We could not evaluate the performance of the syntactic features (SYN) in isolation because they can only be computed for pairs of event mentions from the same sentence, but when combined with EVE and BoW features syntactic features do not seem to have significant impact on relation predictions. Performance for the SUBSUPER class is consistently better than the performance for the SUPERSUB class over all models, indicating that the cases where a *superevent* occurs in text after a *subevent* are somewhat easier to recognize.

5. Conclusion

Event mentions in news denote real-world events of varying spatial and temporal granularity. One event constitutes

	SuperSub			SubSuper			Micro-avg.		
Feature set	P	R	F_1	P	R	F_1	P	R	F_1
EVE	48.1	28.0	35.4	50.5	31.4	38.8	49.4	29.8	37.1
BoW	40.1	26.3	31.8	47.2	26.8	34.2	43.5	26.6	33.0
PS	3.4	39.4	6.3	4.6	1.9	2.7	3.5	20.4	5.9
EVE + BoW	**58.5**	**47.6**	**52.5**	72.0	56.5	63.4	65.2	**52.2**	**58.0**
EVE + SYN	49.7	31.0	38.2	53.2	32.9	40.7	51.5	32.0	39.5
BoW + SYN	44.4	24.8	31.8	48.4	26.0	33.8	46.4	25.4	32.8
EVE + BoW + SYN	**58.5**	**47.6**	**52.5**	72.0	56.6	63.4	65.2	**52.2**	**58.0**
EVE + BoW + PS	58.0	46.3	51.5	**73.0**	**56.8**	**63.9**	**65.5**	51.7	57.8
All	58.4	47.2	52.2	72.8	56.2	63.4	**65.5**	51.8	57.8

Table 1: Pairwise classification performance for automated recognition of relations of spatiotemporal containment

another event only if it is contained by the other event, both spatially and temporally. Extracting event hierarchies based on relations of spatiotemporal containment is important for inference over narratives of news stories. In this work, we created a corpus of manually annotated event hierarchies in news stories to enable computational approaches to the task.

We presented a simple supervised approach for recognizing relations of spatiotemporal containment between events to serve as the baseline for future research. Our initial experiments suggest that recognizing spatiotemporal containment between events is feasible.

In future work we will focus on developing more advanced models for recognizing spatiotemporal containment between events, utilizing event-based knowledge from resources such as WordNet and VerbNet. We will also focus on global approaches for producing spatiotemporally coherent event hierarchies at the document-level by constraining local classification decisions. Apart from global constraints, we will also consider joint learning of spatiotemporal structure at the document level by employing generalized linear learning models (Kordjamshidi and Moens, 2013).

6. Acknowledgments

The first author was supported by the COST STSM grant (Action IC1002, "Multilingual and multifaceted interactive information access"). The third author was supported by the KU Leuven Postdoctoral grant PDMK/13/115.

7. References

ACE. (2005). Evaluation of the detection and recognition of ACE: Entities, values, temporal expressions, relations, and events.

ACE. (2007). The ace 2007 evaluation plan: Evaluation of the detection and recognition of ace entities, values, temporal expressions, relations, and events.

Bethard, S. (2013). ClearTK-TimeML: A minimalist approach to tempeval 2013. In *Second Joint Conference on Lexical and Computational Semantics (*SEM)*, volume 2, pages 10–14, Atlanta, USA. ACL.

Chambers, N. and Jurafsky, D. (2008). Unsupervised learning of narrative event chains. *Proceedings of 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT '08)*, pages 789–797.

Chambers, N. and Jurafsky, D. (2009). Unsupervised learning of narrative schemas and their participants. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 602–610. Association for Computational Linguistics.

Chambers, N. and Jurafsky, D. (2010). A database of narrative schemas. In *LREC*.

Daniel, N., Radev, D., and Allison, T. (2003). Sub-event based multi-document summarization. In *Proceedings of the HLT-NAACL 03 on Text summarization workshop-Volume 5*, pages 9–16. Association for Computational Linguistics.

Fan, R. E., Chang, K., Hsieh, C. J., Wang, X. R., and Lin, C. J. (2008). LibLinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9:1871–1874.

Filatova, E. and Hatzivassiloglou, V. (2004). Event-based extractive summarization. In *Proceedings of ACL Workshop on Summarization*, volume 111.

Glavaš, G. and Šnajder, J. (2013a). Event-centered information retrieval using kernels on event graphs. In *TextGraphs-8 at EMNLP*.

Glavaš, G. and Šnajder, J. (2013b). Exploring coreference uncertainty of generically extracted event mentions. In *Proc. of the CICLing 2013*, pages 408–422. Springer.

Jans, B., Bethard, S., Vulić, I., and Moens, M. F. (2012). Skip n-grams and ranking functions for predicting script events. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 336–344. Association for Computational Linguistics.

Kordjamshidi, P. and Moens, M.-F. (2013). Designing constructive machine learning models based on generalized linear learning techniques. In *Proceedings of the NIPS workshop on constructive machine learning*.

Kordjamshidi, P., van Otterlo, M., and Moens, M. F. (2010). Spatial role labeling: Task definition and annotation scheme. In Calzolari, N., Khalid, C., and

Bente, M., editors, *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC'10)*, pages 413–420.

Mani, I., Hitzeman, J., Richer, J., Harris, D., Quimby, R., and Wellner, B. (2008). SpatialML: Annotation scheme, corpora, and tools. In *LREC*.

Pustejovsky, J., Castano, J., Ingria, R., Sauri, R., Gaizauskas, R., Setzer, A., Katz, G., and Radev, D. (2003a). TimeML: Robust specification of event and temporal expressions in text. *New Directions in Question Answering*, 3:28–34.

Pustejovsky, J., Hanks, P., Sauri, R., See, A., Gaizauskas, R., Setzer, A., Radev, D., Sundheim, B., Day, D., Ferro, L., et al. (2003b). The TimeBank corpus. In *Corpus Linguistics*, volume 2003, page 40.

Quine, W. (1985). Events and reification. *Actions and Events*, pages 162–171.

Roberts, K., Skinner, M. A., and Harabagiu, S. M. (2013). Recognizing spatial containment relations between event mentions. In *10th International Conference on Computational Semantics*.

UzZaman, N. and Allen, J. (2010). TRIPS and TRIOS system for TempEval-2: Extracting temporal information from text. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pages 276–283, Uppsala, Sweden. Association for Computational Linguistics, ACL.

UzZaman, N. and Allen, J. F. (2011). Temporal evaluation. In *ACL (Short Papers)*, pages 351–356.

UzZaman, N., Llorens, H., Derczynski, L., Verhagen, M., Allen, J., and Pustejovsky, J. (2013). Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Proceedings of the 7th International Workshop on Semantic Evaluation (SemEval 2013), in conjunction with the Second Joint Conference on Lexical and Computational Semantcis (* SEM 2013). Association for Computational Linguistics, June*.

Verhagen, M., Gaizauskas, R., Schilder, F., Hepple, M., Katz, G., and Pustejovsky, J. (2007). SemEval-2007 Task 15: TempEval temporal relation identification. In *Proc. of the SemEval 2007*, pages 75–80. Association for Computational Linguistics.

Verhagen, M., Sauri, R., Caselli, T., and Pustejovsky, J. (2010). SemEval-2010 Task 13: TempEval-2. In *Proc. of the SemEval 2010*, pages 57–62.

Wolf, F. and Gibson, E. (2005). Representing discourse coherence: A corpus-based study. *Computational Linguistics*, 31(2):249–287.